View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Open Archive Toulouse Archive Ouverte


Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 12952

The contribution was presented at OM 2013:
http://om2013.ontologymatching.org/

**To cite this version** : Severo, Bernardo and Trojahn, Cassia and Vieira, Renata *PLATAL - a tool for web hierarchies extraction and alignment.* (2013) In: 8th International Workshop on Ontology Matchin (OM 2013), 21 October 2013 - 21 October 2013 (Sydney, Australia).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

# PLATAL - A Tool for Web Hierarchies Extraction and Alignment

Bernardo Severo[1], Cassia Trojahn[2], and Renata Vieira[1]

[1] Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil
[2] Université Toulouse 2 & IRIT, Toulouse, France

**Abstract.** This paper presents PLATAL, a modular and extensible tool for extraction of hierarchical structures from web pages which can be automatically aligned and also manually edited via a graphical interface. Evaluation of alignments can be carried out using standard measures.

## 1 Introduction

Web sites are rich sources of information for a range of applications. Tools for automatically extracting structured content from these sources and for comparing content across web sites are valuable resources. For helping in these tasks, we propose PLATAL (**Plat**form of **Al**ignment), a modular and extensible tool that provides an integrated environment for extraction of web hierarchies and alignment creation, edition and evaluation. The main motivation behind PLATAL is to assist users in the complete alignment cycle of two web hierarchies. Differently from other matching tools offering a visual environment, like OLA [1], Prompt [3], Homer [5], Yam++ [2] and SOA-based tool [4], PLATAL offers novel functionalities: the possibility of automatically extracting hierarchical structures from the web together with a centralised visual tool for alignment manipulation.

## 2 PLATAL modules

PLATAL is a standalone tool composed of four modules: (1) *hierarchy extraction module*, which extracts fragments from HTML pages using XPath expressions; (2) *automatic alignment module*, which implements a set of terminological (prefix, suffix, edit-distance) and structural matching techniques (similarity of parents and children entities) for generating equivalence correspondences; (3) *manual alignment module*, which allows users to edit or create alignments; and (4) *evaluation module*, which takes two alignments and computes precision, recall and F-measure measures. These modules operate independently of each other and alternative implementations can be added instead. Figure 1 shows a screenshot of automatic alignment creation. After loading two hierarchies, each hierarchy will be displayed in the respective section. Then, users can select one or more alignment processes and start them ('Start Alignment Process'). If at least one method founds one correspondence between two entities, the user can see it by

selecting the source or target entity in the hierarchies (field 'Correspondences'). Alignments can be exported in the Alignment format[3] ('Save').

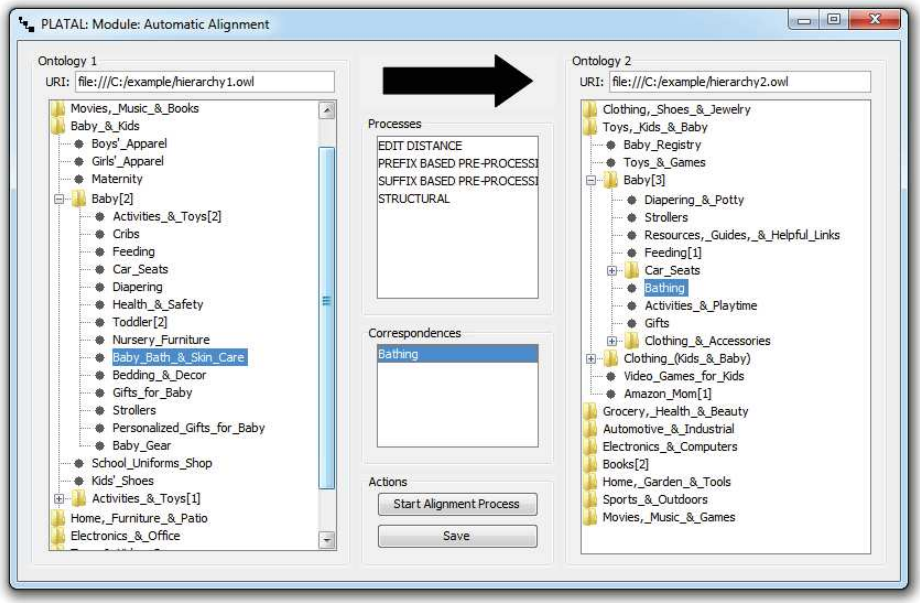


**Fig. 1.** Automatic Alignment Module screenshot.

## 3 Conclusions and future work

We have presented a visual tool for extraction, alignment and evaluation of web hierarchies. To the best of our knowledge, there is no publicly available environment integrating all these features together. As future work, we plan to improve the visualisation of alignments, develop a web-based version, allow parametrisation and customisation of alignment techniques through the user interface, and add a multilingual ontology matching module.

## References

1. J. Euzenat, D. Loup, M. Touzani, and P. Valtchev. Ontology Alignment with OLA. In *3rd EON Workshop*, pages 59–68, 2004.
2. D. H. Ngo and Z. Bellahsene. YAM++ : (not) Yet Another Matcher for Ontology Matching Task. In *BDA*, France, 2012.
3. N. F. Noy and M. A. Musen. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. In *17th AAAI*, pages 450–455, 2000.
4. K. W. Onn, V. Sabol, M. Granitzer, W. Kienreich, and D. Lukose. A visual soa-based ontology alignment tool. In *OM*, 2011.
5. O. Udrea, R. Miller, and L. Getoor. Homer: Ontology visualization and analysis. In *Demo session ISWC*, 2007.

[3] http://alignapi.gforge.inria.fr/format.html